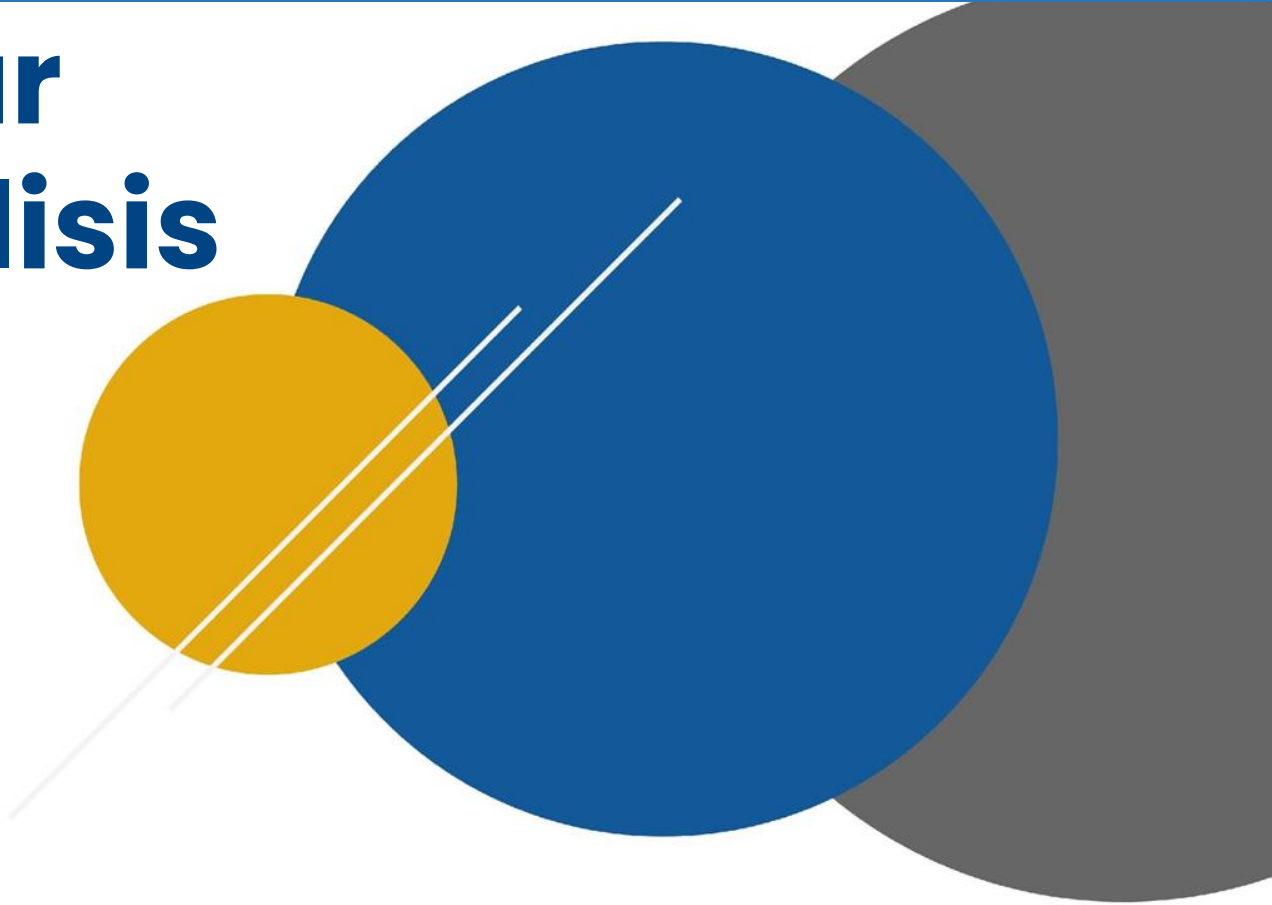


Pengantar Data Analisis





Apa yang akan dipelajari?

- Pengenalan *Data Science*
- Peluang dan Tantangan
- Konsep CRISP-DM (*Cross Industry Standard Process for Data Mining*)

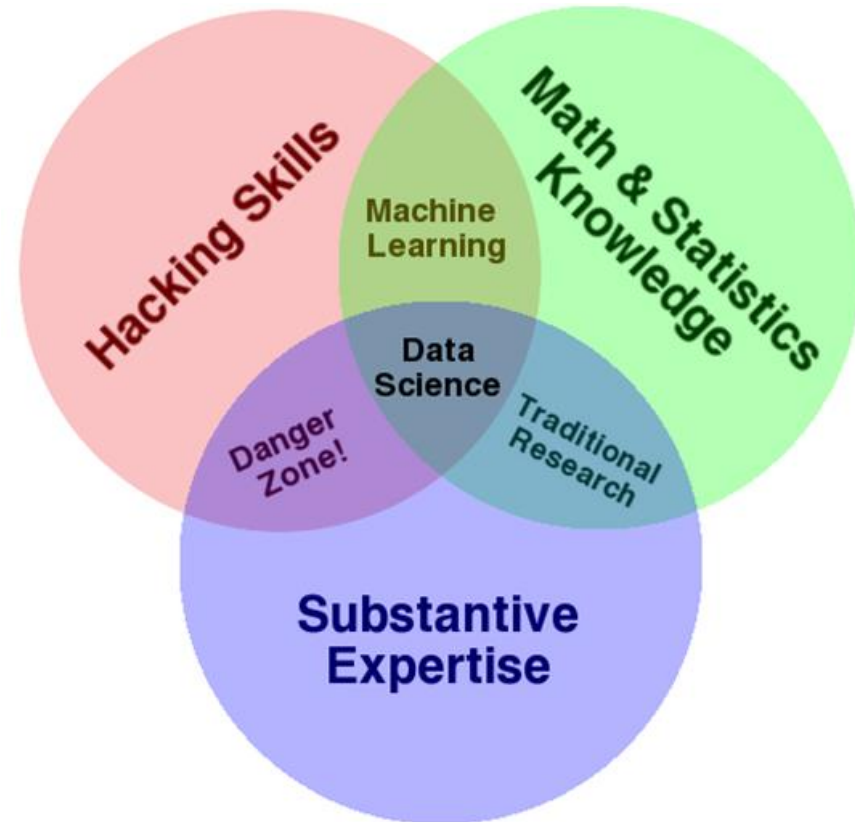




Apa itu *Data Science*?

Data Science adalah bidang dengan multidisiplin yang menggunakan metode *scientific*, proses, algoritma dan sistem yang mengekstraksi pengetahuan dari data yang terstruktur maupun tidak. keterampilan yang diperlukan diantaranya *Statistics*, *Computer Science* dan *Domain Expertise*.

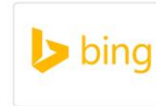
1. *Statistics* untuk memodelkan dan meringkas kumpulan data
2. *Computer Science* untuk merancang dan menggunakan algoritma untuk menyimpan, memproses, dan memvisualisasikan data
3. *Domain Expertise*, diperlukan merumuskan pertanyaan dan dapat menempatkan jawaban dalam konteks





Penerapan *Data Science*

- Security
- Internet Search
- Fraud and Risk Detection
- Banking dan Finance
- Digital Advertisements
- Recommender System
- Image Processing
- Speech Recognition
- Price Comparison Websites
- Airline Routing Planning
- Delivery Logistics
- Health Care
- Sport/ Gaming
- ect.



DATA SECURITY





Kenapa *Data Science*?

Gap antara pertumbuhan Data terhadap pertumbuhan Data Analysts

-Big Data, Sejak semua berbasis internet, data menjadi kunci penting perusahaan berbagai industri. Setiap tindakan pengguna secara online ditelusuri dan jejaknya dapat menyebabkan data dalam jumlah yang besar.

High Demand for Data Scientist

-Artificial Intelligence, karena meningkatnya data online, permintaan data science meningkat sehubungan dengan teknologi baru, AI.

Data Scientist job openings at the world's top companies



Data from Thinknum - Open dataset

• Title (Count)

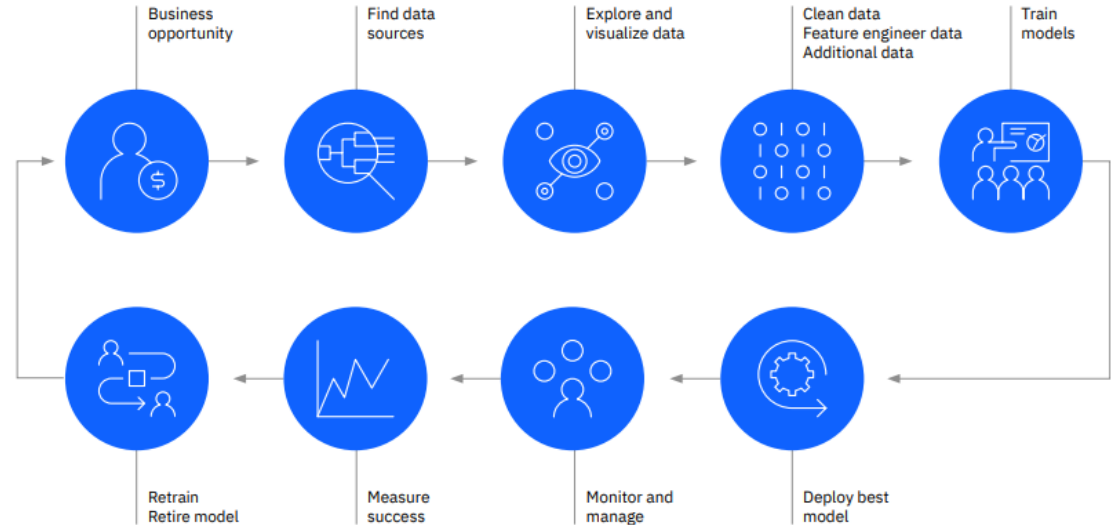
Membutuhkan 28% lagi Data Scientist pada tahun 2020 untuk bisa mencapai pertumbuhan Data





Keterlibatan *Data Science* dalam AI

- Memahami peluang bisnis
- Menemukan sumber data yang tepat
- Menyiapkan data dan membangun Machine Learning dan model khusus AI
- Membantu penyebaran model dalam organisasi
- Menjadi alat ukur kesuksesan dan komunikasi bisnis



AI Workflow





Keterampilan Dasar *Data Science*

- Dasar Statistika dan Pemograman
- Dasar *Data Science*
- *Data Preparation* (Identifikasi, Pengumpulan data, Memanipulasi, mengubah dan membersihkan data)
- *Model Building*
- *Model Deployment*
- *Big Data Foundation*
- Kepemimpinan dan Perngembangan yang Profesional





Tantangan *Data Science*

Data Science, tidak mudah karena memiliki beberapa tantangan dengan kualitas dari aspek *Context, Data, Process, Result*

- **The Complexity of Reality** (raw data, misunderstood, value of knowledge)
- **Modelling process** (aware terhadap kompleksitas dan mengerti proses)
- **Cognitive Bias** (human bias, interpretation, possible subjective alterations)
- **Kualitas Data** (Akurasi, Kelengkapan, Data Unik, *Up-to-date* Data dan Konsistensi) -> *Vicious Cycle of Bad Data*
- Sample dan Data dari sumber berbeda, apakah data yang representatif?
- Penyeragaman data
- Ketersediaan dan kemudahan akses data





Peluang *Data Science*

- Memahami peluang bisnis dengan pengetahuan data yang tepat
- Menjadi model, alat ukur kesuksesan dan komunikasi bisnis
- Membangun Machine Learning dan model khusus AI
- Penerapan diberbagai aspek





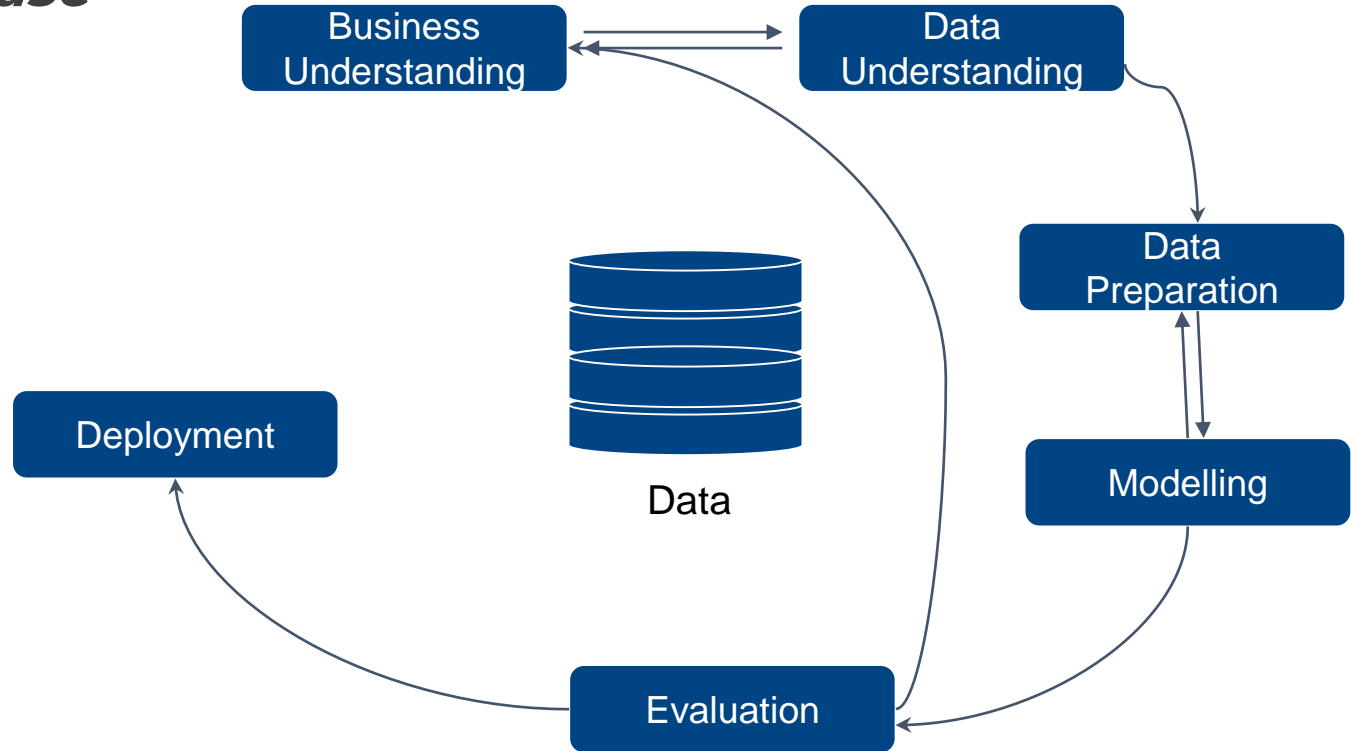
Data Analysis

- ***Data Analysis*** adalah proses dari pemeriksaan, pembersihan, transformasi dan pemodelan data dengan tujuan untuk menemukan informasi yang berguna, mendapatkan wawasan yang dapat ditindaklanjuti, mendukung pengambilan keputusan dan memberikan kesimpulan.
- Beberapa teknik analisa termasuk, ***Data Mining, Bussiness Intelligence, Descriptive Statistics (Univariate analysis, Bivariate analysis, Multivariate analysis)***





CRISP-DM Phase





Outline CRISP-DM Phase

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives <i>Background Business Objectives Business Success Criteria</i>	Collect Initial Data <i>Initial Data Collection Report</i>	Select Data <i>Rationale for Inclusion/ Exclusion</i>	Select Modeling Techniques <i>Modeling Technique Modeling Assumptions</i>	Evaluate Results <i>Assessment of Data Mining Results w.r.t. Business Success Criteria Approved Models</i>	Plan Deployment <i>Deployment Plan</i>
Assess Situation <i>Inventory of Resources Requirements, Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits</i>	Describe Data <i>Data Description Report</i>	Clean Data <i>Data Cleaning Report</i>	Generate Test Design <i>Test Design</i>	Review Process <i>Review of Process</i>	Plan Monitoring and Maintenance <i>Monitoring and Maintenance Plan</i>
Determine Data Mining Goals <i>Data Mining Goals Data Mining Success Criteria</i>	Explore Data <i>Data Exploration Report</i>	Construct Data <i>Derived Attributes Generated Records</i>	Build Model <i>Parameter Settings Models Model Descriptions</i>	Determine Next Steps <i>List of Possible Actions Decision</i>	Produce Final Report <i>Final Report Final Presentation</i>
Produce Project Plan <i>Project Plan Initial Assessment of Tools and Techniques</i>	Verify Data Quality <i>Data Quality Report</i>	Integrate Data <i>Merged Data</i>	Assess Model <i>Model Assessment Revised Parameter Settings</i>		Review Project <i>Experience Documentation</i>
		Format Data <i>Reformatted Data</i>			
		<i>Dataset Dataset Description</i>			





What's Next?

Pengenalan Bahasa Pemrograman Python

PENGENALAN DATA ANALISIS

